

## AN ANECHOIC AUDIO CORPUS FOR ROOM ACOUSTICS AND RELATED STUDIES

*Antti Kuusinen*

Virtual Acoustics Research Group, Department of Media Technology,  
Aalto University School of Science,  
Espoo, Finland  
antti.kuusinen@aalto.fi

### ABSTRACT

Anechoic or semi-anechoic instrument recordings are readily available for academic purposes on a few different sites online. Anechoic recordings are commonly used in auralizations, which today practically means convolving recordings with simulated or measured room impulse responses. Besides the possibility of being used as such, these recordings offer other possibilities for the generation of test stimuli. Many studies, such as, studies on auditory distance perception or source separation, would benefit from available experimental materials which would not be strictly musical but could still be linked to the perception of musical stimuli. The goal of the current investigation is to develop a procedure for generating such materials, i.e., an anechoic audio corpus which can be used in the future investigations of room acoustics and in related fields. Moreover, the aim is to provide a framework for further development of processes where a large number of stimuli can be generated in a systematic way. In this study, the proposed framework is instantiated by producing two sets of stimuli by either directly segmenting anechoic music or randomly combining different segments of anechoic instrument tracks. Music information retrieval (MIR) approach is used to calculate 14 musical features of the generated sets of stimuli. Principal component analysis is used to analyse the sample spaces enabling the experimenter to select a small number stimuli with desired characteristics. The benefits and drawbacks of this stimuli generation approach including some important theoretical underpinnings of experimental design are also discussed.

### 1. INTRODUCTION

Auralizations are made by convolving audio signals with simulated or measured impulse responses. Making convolutions commonly require anechoic audio materials which are free from extraneous reflected sounds - especially in research of room acoustics. In the studies of music performance spaces, anechoic instrument recordings or synthesized instrument sounds are natural choices as source signals. Anechoic or semi-anechoic instrument recordings are readily available for academic purposes on a few different sites online, e.g., [1, 2]. Also commercial releases, such as, Vienna Symphonic Library [3] exist. Such recordings are usually of either single notes with various characteristics (steady, vibrato, pizzicato etc.) and/or recordings of musical sequences or excerpts of compositions with various styles and with separate tracks for each instrument or instrument section [4].

Besides being invaluable for research as such, these recordings offer other possibilities for the generation of stimuli. Many studies, such as, studies on auditory distance perception or source separation, would benefit from available experimental materials

which would not be strictly musical but could still be linked to the perception of musical stimuli. Also studies where 'ecologically valid' musical stimuli, that is, stimuli which are unequivocally musical are required, could benefit from contrasting stimuli which would not be strictly musical. While the ecological validity of the test stimuli is one of the main reasons for employing commercial recordings and well known compositions, usually the recording and production chain is not well described, what is a major drawback to using such releases in scientific work. In addition, there are many situations where the selection of test stimuli unwarrantedly restricts the experimenter from making more generalizable inferences from the results. This is true especially in situations where musical stimuli are used in the subjective evaluation of "treatments" such as different signal processing algorithms, auralization methods, or different room acoustical conditions.

Consider a simple case where a researcher investigates the performance of a few reverberation algorithms with a listening experiment. Let's say that the number of algorithms, i.e., treatments, to evaluate is 6. The perceptual task is to indicate for each sound how "natural" the reverberation is on a 20-point scale. Then he or she chooses four anechoic source signals with different characteristics which he/she thinks is a representative sample of the population of sounds to which the algorithms would be used. Listening experiment is conducted in blocks, where the algorithms are compared in parallel with each source signal. A number of assessors participates in the experiment and data is analysed with the analysis of variance (anova).

The different sources of variance, that is, factors, must be specified in setting up an anova model. Moreover, one must specify whether each factor is treated as "fixed" or "random". In our example, the main factors are the treatments (the algorithms), the source signals and the individuals. Whether to treat these factors as fixed or random is essentially based on the nature of inferences one wishes make to about the possible results. Treating a factor as fixed indicates that the different levels of that factor are exhaustive of the population that the factor represent, that is, the inferences are restricted to these particular treatments. Clearly, the treatments are considered as fixed in our current example as the objective is to evaluate the perceptual differences in this particular set of algorithms. In contrast, treating a factor as random indicates that the factor levels are drawn randomly from the population of interest and possible inferences made from the results generalize to that population. In other words, the differences in these levels of the factor are not of particular interest, but rather the effect of changes between the factor levels in general. In our example, individuals can be considered as being randomly selected from a population and thus, as a random factor.

Treating the algorithms as a fixed factor and the individuals

as a random factor is quite straightforward, but the source signal might be considered as a fixed or a random depending on the assumptions and inferences the experimenter wishes to make. If treated as a fixed factor, the conclusions are made specifically about these source signals. In other terms, the algorithms are studied only with respect to these particular source signals making the experiment into a 'case' study. It is of course possible, that as a case study, the possible inferences may have apparent theoretical implications which account for a more general discussion about the results. In addition, treating a factor as fixed also has the benefit for statistical analysis to reveal smaller effects, which is apparent by considering the sources of variances included in the error component of a model. A detailed discussion about these aspects is outside the scope of this paper (see more in, e.g., [5] or [6]), but basically treating a blocking factor, in our case the source signal, as random, adds uncertainty (i.e., wider confidence intervals) to the analysis of treatment means. However, if the treatments are still significantly different from each other when the source signal is treated as a random factor, the inferences from these results can be generalized to the population from which these source signals are randomly chosen.

Thus, in our example, treating the source signal as a random factor means that the experimenter can generalize the possible results about the performance of the reverberation algorithms outside the set of source signals used in the experiment. This is clearly desirable in many situations, but, this also implies that the source signals should be a representative sample randomly drawn from a population of possible source signals. Clearly, this population can be stated as infinite and impossible to specify, so that the experimenter may well choose the source samples on the grounds of his/her best knowledge and intuition. Unfortunately, subjective knowledge and intuition are often rather problematic bases for scientific work. Thus, here I propose a simple way to help researchers to select source signals by first producing a large set of samples, that is a population of stimuli, which can then be randomly sampled as desired. Of course, the issue of the representativeness of this population to an infinite sound space still remains, but the major advantage is the knowledge about the population of stimuli to which the results should be generalizable. Keeping these considerations in mind, the next section provides some further contextual aspects and reasons for why such processes are needed.

## 2. SELECTING STIMULI FOR LISTENING EXPERIMENTS

As discussed above, the information the experimenter expects to extract from the results determines the design of the experiment - including the selection of the test stimuli. In different fields of audio research different types of stimuli are typically used. For instance, in psychoacoustic research often used stimuli are noises and pure tones with various derivatives and variations (see e.g. [7], p. 2). In speech perception studies a natural choice is to use samples of speech. In music related studies some excerpts of music is a common choice. The performances of reproduction systems and perceptual coding algorithms are commonly studied with stimuli which is known to be particularly revealing about a certain effect; e.g., sound of castanets is typically used to reveal about unwanted pre-echos in perceptual coding. A simplified schematic of different stimulus types is presented in Fig. 1. Of course any combination of different types of stimuli is possible if needed and in fact, often there is no clear distinction between stimulus types. In Fig. 1

the miscellaneous stimuli are presented by a big surrounding circle while the more specific types of stimuli are represented by the circles.

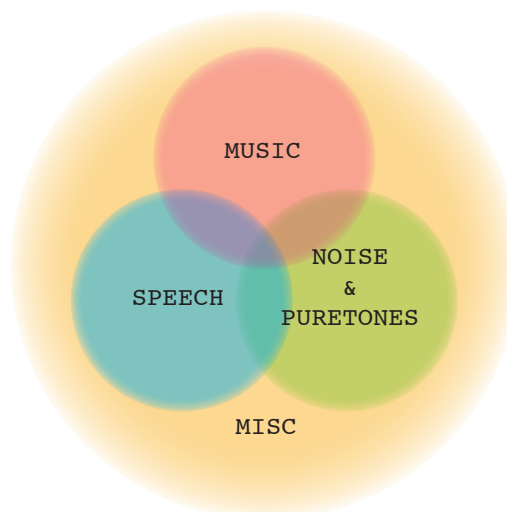


Figure 1: A schematic of different types of stimuli.

There are many fields of audio research, such as room acoustics, where practically any stimulus type can be used. The generation and manipulation of noises and pure tones have become everyday practice with modern digital signal processing techniques and now these types of stimuli can be quite easily obtained when needed. Considering speech related studies, the situation is not as straightforward, but there are readily available speech samples which have been widely employed in speech intelligibility and related studies. The Coordinate Response Measure (CRM) corpus [8] was originally developed for a particular speech recognition task but has been extended [9] to enable studies also in audio-visual domain. The many advantages of a stimulus set, which is widely adopted and used in a research field, include: facilitated experimental design, easier comparison and cross-validation of studies performed in different laboratories and possibly more efficient "steering" of research with the emergence of relevant research problems.

Studies of music and related aspects are performed in a wide variety of research fields ranging from neuroscience, psychology, aesthetics and musicology to recording techniques, signal processing and room acoustics. Test stimuli are commonly excerpts of music which are most often selected by the researcher. Regarding research in music and emotion, for instance, Eerola et al. [10] reviewed 170 studies and report that the stimuli selection method has been almost entirely researcher-driven (96 %) in a sense that the choices were based on either researcher's intuition (33 %), on a pilot study (8 %), on a selection by a group of experts (6 %) or on some previous study (9 %). Although these numbers might be relevant only in that particular field, there is little doubt that researchers' knowledge and intuition would be the main determinants in the selection of test material in most experiments where musical stimuli are required, independent of the research field. Moreover, in the field of neuroscience the lack of a systematic approach to selecting musical stimuli has been argued to be one possible reason for inconsistencies between studies [11].

Based on this discussion and a few statistical principles mentioned before, it is clear that systematic approaches to stimulus selection would be beneficial for various fields of audio research. Here, one such method is proposed to help researchers in the sample selection process, as well as to stipulate critical discussion on the topic.

### 3. ANECHOIC AUDIO CORPUS

In order to enable a random sampling of source stimuli and to provide a framework for sample selection and future developments of similar processes, an audio corpus and a simple stimuli production method is proposed. The stimuli production method is based on the segmentation of pre-recorded anechoic material and making combinations of these segments. In particular, the aim of the current investigation is to generate a population of test materials which would allow random sampling, and where the stimuli 1) would be consisted of anechoic sounds of real instruments. 2) would include both musical stimuli as well as stimuli which are not distinctively musical as in the terms of a melody or a harmony, 3) would enable control over the acoustic ("musical") features [12] (dynamics, rhythm, timbre, pitch and tonality) of the selected stimuli and 4) could be used in a wide range of listening experiments. Moreover, the aim is to develop a systematic stimuli selection procedure, which can be used in conjunction with the experimenter's knowledge and intuition. It is worth to mention that the current work does not attempt to produce a selection procedure for "ecologically valid" musical signals although the first method proposed below also fulfills this criterium to a certain extent. The aim is to provide a framework for further development of processes where (random) stimuli can be generated and selected in a systematic way.

#### 3.1. Background

The starting point for the current work is the anechoic symphony orchestra recordings made by Pätynen *et al.* [4]. These recordings consist of the following excerpts:

- W. A. Mozart: Aria of Donna Elvira from the opera Don Giovanni (3 min 47 s)
- L. v. Beethoven: Symphony no. 7, I movement, bars 1-53 (3 min 11 s)
- A. Bruckner: Symphony no. 8, II movement, bars 1-61 (1 min 27 s)
- G. Mahler: Symphony no. 1, IV movement, bars 1-72 (2 min 12 s)

The Fig. 2 represents the chromagrams of the music pieces. Each instrument has been recorded (48 kHz, 16 bits) separately in an anechoic chamber. Also some editing and noise reduction have been applied on the separate instrument tracks (see details of the recording process in [4] and [13]). Clearly, one could also use any recordings, such as, single notes, but in order to preserve musical characteristics, which occur naturally in played music, such as transitions, these recordings were thought to be the most appropriate for the current investigation.

The characterization of the stimulus space is inspired by recent developments in the field of music information retrieval (MIR), where acoustic features calculated directly from audio signals are used for various purposes, particularly for automatic classification

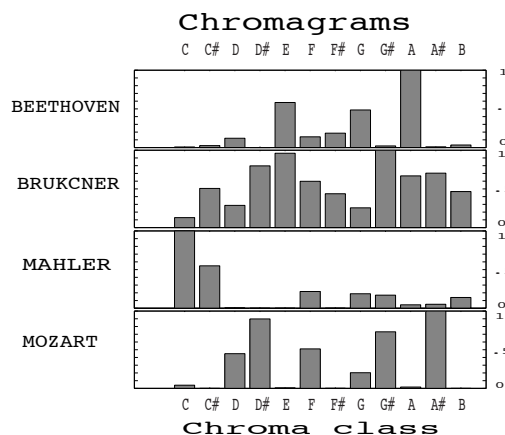


Figure 2: Chromagrams of the original excerpts of symphonic music.

tasks (genre, mood etc.) (see e.g. [14] for review). The extracted features are thought to represent the essential musical characteristics in the signal and to correspond to human perception to some extent. Commonly, the features represent the musical dimensions of pitch, dynamics, rhythm, timbre and tonality. Previously feature extraction and classifier training were performed in Marsyas [15] framework with WEKA machine learning software but more recently also a Matlab toolbox (MIRtoolbox [12]) has been developed. Matlab and MIRtoolbox are used in the current investigation for acoustic feature extraction. Otherwise, the data analysis is performed in R statistical programming language.

#### 3.2. Acoustic features

A set of 14 acoustic features were selected to characterize the musical properties of the stimuli. Regarding the computational time needed for calculating each feature for each individual sample, the following features were considered covering the main musical aspects including timbre, dynamics, tonality and rhythm:

- Timbral:
  - zero-crossing rate (sign change in the signal per second),
  - spectral roll-off (a cut-off frequency below which lies 85 % of total energy),
  - brightness (energy above 1500 Hz),
  - spectral flatness,
  - spectral entropy,
  - roughness.
- Dynamics:
  - root-mean-square energy (frame length of 50 ms, 50 % overlap)
  - low to high energy ratio (% of frames with less than average energy in the segment)
- Tonality:
  - spectral flux (distance between the spectrum of successive frames of 50 ms, 50 % overlap),
  - key clarity,
  - mode (minor (-1) – major (+1)),
- Rhythm:
  - Tempo

The values were averaged over the duration of each segment to obtain single values for subsequent analysis. The calculation of these features was performed by using the default configurations defined in MIRtoolbox [12] because there was no reasons were found to change these settings. However, it is acknowledged that the settings, such as, the durations of temporal windows and the percentages of overlap affect the calculations, and closer investigation to these aspects would be beneficial for future work.

### 3.3. Stimulus production

To make the stimulus production process more tangible, the length of the stimuli to produce was arbitrarily set to five seconds. Of course, in real applications the length of the stimuli would be determined by the experimental design and the context and objectives of a study. Two different stimulus production methods are proposed. The first method produces "ecologically valid" musical stimuli, but is restricted by the available musical variations in the original music pieces as well as the desired length of stimuli. The second method produces a larger set of random stimuli, which might be musically questionable but enables control over various other aspects of the stimulus set, such as the number of instruments. Both of these methods enable a random sampling of the produced stimulus population. In addition, the characterization of these stimulus populations by musical features combined with principal component analysis allows one to include covariates in statistical models, thus enabling the evaluation of these effects in the results. The principal component analysis used below is performed on the correlation matrix.

#### 3.3.1. Method 1.

First method is inspired by a method represented by Alluri *et al.* [16]. The aim is to select five second segments from the original music pieces, that would capture the range of musical variations embedded in these music pieces. The original instrument recordings are combined into one channel and 5 second segments with 1 second hop size (80 % overlap) are extracted. Including all music pieces, the total number of segments obtained and subsequently analyzed was 621. Musical features are calculated for each segment and the features containing values for shorter temporal windows are averaged over the 5 second period. Then, principal component analysis (PCA) is performed in order to reduce the dimensionality of the feature space, and to reveal associations between different features. Principal components with eigenvalues larger than 1 are selected, and varimax rotation is performed with the selected components in order to clarify the structure and interpretation of the reduced space. Finally, this reduced space can be used to select sets of samples in various ways.

Here, the averaged values of the features of the five second segments from all recordings were combined to the same data matrix which was subjected to PCA. The first 5 principal components explained 81 % of the total variance and were retained. Table 1 presents the feature loadings these PCs after varimax rotation. Although the interpretation of principal components should be validated by a perceptual experiment, as in Alluri *et al.* [16], the PC loadings indicate that the first PC is associated with the timbral properties, the second to the sensory dissonance or consonance, the third to the dynamics and PC4 and PC5 to the tonal characteristics of the samples.

Table 1: Feature loadings on five principal components after varimax rotation. Features were calculated for 621 stimuli produced by extracting 5 second segments of anechoic music (see text for details).

% of var.	PC1 32 %	PC2 14 %	PC3 13 %	PC4 13 %	PC5 13 %
Zerocross	<b>0.74</b>	-0.09	0.37	-0.04	-0.12
Rolloff	<b>0.77</b>	0.42	0.14	0.31	0.10
Brightness	<b>0.89</b>	-0.05	-0.02	-0.07	-0.19
Flatness	<b>0.80</b>	0.51	0.05	0.20	0.14
spcEnt	<b>0.91</b>	-0.17	-0.06	-0.02	0.13
spcCent	<b>0.85</b>	0.40	0.17	0.17	0.05
Roughness	0.07	<b>-0.87</b>	0.00	0.21	0.16
Spread	0.52	<b>0.71</b>	0.04	0.34	0.20
Rms	-0.31	0.10	<b>-0.85</b>	-0.02	0.16
Lowenergy	-0.02	0.14	<b>0.83</b>	-0.07	0.05
spcFlux	-0.11	0.05	-0.39	<b>-0.80</b>	0.06
Keycla	0.01	-0.01	-0.18	<b>0.79</b>	0.16
Mode	0.02	-0.04	-0.06	0.00	<b>0.92</b>
Tempo	-0.05	-0.03	0.32	-0.47	0.24

#### 3.3.2. Method 2

In contrast to the method described above, here the aim was to produce a large set of stimuli, which consist of anechoic instrument sounds but are not predetermined or deliberately composed in musical terms. First, each anechoic instrument recording is segmented at silent periods determined by root-mean-square energy in 500 ms frames and 100 ms hop size. The segments from each music piece were grouped into instrument banks. In the present case 15 different instrument banks were obtained, reflecting the instrumentation of the original compositions. The respective instruments are bassoon, cello, clarinet, contrabass, French horn, flute, oboe, timpani, trombone, trumpet, tuba, viola, 1. violins and 2. violins.

To produce a large randomized set of 5 second samples, one sample of each instrument bank was randomly selected and combined with randomly selected samples of other instrument banks. Here, only segments shorter than 5 seconds were used. The temporal positions of the selected segments were randomly varied inside the five second sample to avoid 'stacking' the sounds to the beginning of the samples.

In this randomization procedure, it is also possible to control the number of instruments included in the random combinations. In the current implementation, the composition of instruments in a classical orchestra was used with an added trombone and a percussion instrument. In this orchestration the number of instruments, i.e., the number of randomly selected samples of the instrument banks was as follows: 2 flutes, 2 oboes, 2 clarinets, 2 bassoons, 2 French horns, 2 trumpets, 2 timpani, 10 1. violins, 8 2. violins, 6 violas, 4 cellos and 2 double bass. 1000 random combinations of the segmented instrument samples in this orchestration were produced. Features were calculated for each combination separately and average values were extracted. Like in the previous method, this feature set was analysed using PCA and varimax. Again, the first 5 principal components were retained and together they explained 82 % of the total variance. However, in contrast to the first method where the explained variances were more equally dis-

tributed between the PCs, we now observe that the PC1 explains as much as 43 % of the total variance even when the varimax rotation tends to make the explained variances more equal. The feature loadings on these PCs are tabulated in Table 2. The loadings indicate that again the first component is associated to timbral features, but now also dissonance related aspects are included. The second component is associated with dynamics, the third with tonality, and interestingly mode and tempo parameters uniquely characterize the fourth and fifth component, respectively.

Table 2: Feature loadings on five principal components after varimax rotation. Features were calculated for 1000 stimuli produced by randomly combining the 5 second segments of anechoic instrument sounds (see text for details).

% of var.	PC1 43 %	PC2 15 %	PC3 9 %	PC4 7 %	PC5 7 %
Zerocross	<b>0.79</b>	0.22	-0.13	0.03	-0.04
Rolloff	<b>0.93</b>	-0.11	-0.04	-0.04	0.00
Brightness	<b>0.87</b>	0.16	-0.17	0.01	-0.05
Flatness	<b>0.93</b>	-0.28	0.02	-0.05	-0.03
spcEnt	<b>0.93</b>	-0.03	0.13	0.01	-0.03
spcCent	<b>0.98</b>	-0.06	-0.08	-0.04	-0.03
Spread	<b>0.69</b>	<b>-0.53</b>	-0.08	-0.07	-0.01
Roughness	<b>0.70</b>	0.29	0.14	0.11	0.05
Rms	-0.02	<b>0.90</b>	0.06	0.00	0.00
Lowenergy	-0.05	<b>-0.82</b>	0.12	0.03	-0.01
spcFlux	0.11	0.10	<b>0.91</b>	-0.03	-0.02
Keycla	0.31	0.29	<b>-0.58</b>	-0.16	-0.14
Mode	0.01	-0.02	0.05	<b>0.99</b>	0.01
Tempo	-0.03	0.01	0.05	0.01	<b>0.99</b>

### 3.4. Selection of samples

The sample space characterized by musical features and ordinated by PCA with varimax rotation can be used to select samples in various ways. However, the sample scores on the principal components could be used to constrict the random sampling on some subpopulations of interest. For example, if an experiment requires that the samples should not be very dissimilar in terms of dynamics, one could calculate a percentile (e.g., 25th) cutoff scores and make random sampling only to the subpopulation inside that range. Otherwise, for instance a clustering analysis could provide interesting possibilities for sample selection where one could choose samples which are very similar or dissimilar with respect to this sample space characterization. Also the rank ordering of samples scores combined with equidistant sampling would provide sets of samples where the variations between the samples in each set would represent the ranges of variations in each component. This sample selection method could also be used to perceptually validate the interpretation of principal components as shown by Alluri *et al.* [16]. The perceptual validation of the principal component spaces presented in the current investigation is left for future work but is acknowledged to be an important step in the development of this audio corpus and the proposed stimulus selection framework.

## 4. DISCUSSION

Anechoic recordings are continuously used in auralizations in room acoustics and related fields. While in many studies the stimuli have been successfully selected by relying on researchers' expert knowledge and intuition, such practice is susceptible to experimenter bias and problematic for scientific work. Here, a framework for stimulus production and selection procedure was developed to alleviate this issue, but the applicability of this work remains to be validated in practice. The proposed stimulus production method takes advantage of the possibility to automatically segment (and combine) anechoic instrument recordings of symphonic music. The resulted anechoic audio corpus consists of a large number of short segments of anechoic instrument sounds. The segments contain not only individual notes but also short passages and transitions between notes. Sounds of 15 different instruments of a symphony orchestra are currently included in the corpus which is available online for academic purposes.

Only four short pieces of symphonic music were employed in the production of sound samples, what evidently restricts the representativeness of the audio corpus at the moment. Nevertheless, the proposed framework for a stimulus selection procedure is not restricted to only this sample space but can be implemented in a wide range of studies where the experiment is not bound to a specific type of stimuli. Although the issue of generalizability and representativeness of the test stimuli may still remain even when the proposed approach is advocated, it is already a major advantage to have an explicitly described systematic approach as a stimuli selection procedure, instead of just relying on intuition and subjective opinion. Also the characterization of the stimulus space enables the experimenter to make experimental designs and corresponding statistical models where the influence of the properties of the anechoic stimuli can be analysed. This approach can be also used as a tool to guide the experimenter, even though the final selection is performed on a subjective basis.

In the current work, the proposed framework was used to produce two large sets of 5 second long sound samples from four short pieces of anechoic symphonic music. The first set consisted of segments of music where the music was left as composed - albeit cut from the context due to the desired length of stimuli. Clearly, the length of the stimuli used here did not allow for the evaluation of numerous intrinsic and essential aspects of music, such as, longer phrases, verses, choruses, motifs, harmonic developments etc. Such compositional properties of music were not targeted in the current work, where the focus was on lower level acoustic properties in the signals. Extending the length of the extracted segments would be straightforward and would also allow for analysing a number of higher level structures although these structures would be effectively restricted by the musical material. An interesting alternative would be to randomly concatenate the segments of individual instruments to produce random "streams" of instrument sounds, which could be in turn combined with other instrument streams. Such random "music" could be used as a contrastive stimulus to be compared with composed music and for other explorative investigations of higher level musical percepts.

A complementary procedure and a second stimulus set was produced by segmenting the separate instrument recordings and randomly combining these segmented parts. This way also the number of instruments in the produced stimuli could be controlled although this option was not exploited in the current work. Again, the length of the stimuli was fixed to 5 seconds, but variable length

stimuli could also be easily produced. Considering the second method which produces stimuli which are not distinctively musical, calculating features designed to capture perceptually relevant musical features is ambiguous. Other issues which should be addressed in the future work are, for instance, the effect of the window sizes used in calculating the features and determining the most relevant features which would capture the most essential properties of the anechoic signals. The feature set used in the current implementation was limited to 14 features, excluding features, such as, mel-frequency-cepstral coefficient (and its derivatives), pulse clarity, fluctuation centroid and fluctuation entropy. A closer look at these and other features will be taken in the future.

In sophisticated auralization schemes, such as the ones used for studying concert hall acoustics [17], and with appropriate experimental design, this framework could be used to reveal systematic dependencies between the room acoustical properties and the musical features of the source signals. In addition one could also analyze the influence of the instrumentation in the orchestra on the musical features and/or perception of auralization or other signal processing algorithms.

At the moment, the value of the proposed stimuli selection procedure is theoretical and it remains to be studied if in the same experiment the stimuli produced and selected by this procedure will result in a significantly different outcome than the stimuli hand-picked by the experimenter. Researchers are encouraged to explore these and other possibilities as the pre-processed and segmented instrument files and the full length recordings are freely available online. In addition, some basic Matlab scripts for the segmentation and feature extraction are available by request, but potential users are strongly encouraged to write their own scripts as MIRtoolbox is well documented and provided with an extensive user manual.

## 5. ACKNOWLEDGMENTS

Thanks to Tapio Lokki for discussions about the topic and valuable remarks on this manuscript. Academy of Finland [257099] is thanked for financial support. Also many thanks to the great number of anonymous reviewers!

## 6. REFERENCES

- [1] "University of Ferrara," Available at <http://acustica.ing.unife.it/eng-ver/ricerche-eng/Architectural.html>, accessed October 30, 2013.
- [2] "University of Iowa Electronic Music Studios," Available at <http://theremin.music.uiowa.edu/MIS.html>, accessed October 30, 2013.
- [3] "Vienna Symphonic Library," Available at <http://vsl.co.at/>, accessed October 30, 2013.
- [4] J. Pätynen, V. Pulkki, and T. Lokki, "Anechoic recording system for symphony orchestra," *Acta Acustica united with Acustica*, vol. 94, no. 6, pp. 856–865, 2008.
- [5] R. H. Baayen, D. J. Davidson, and D. M. Bates, "Mixed-effects modeling with crossed random effects for subjects and items," *Journal of memory and language*, vol. 59, no. 4, pp. 390–412, 2008.
- [6] A. Gelman and J. Hill, *Data analysis using regression and multilevel/hierarchical models*, Cambridge University Press, 2007.
- [7] H. Fastl and E. Zwicker, *Psychoacoustics: facts and models*, vol. 22, Springer, 2007.
- [8] R. S. Bolia, W. T. Nelson, M. A. Ericson, and B. D. Simpson, "A speech corpus for multitaler communications research," *The Journal of the Acoustical Society of America*, vol. 107, pp. 1065–1066, 2000.
- [9] M. Cooke, Jon Barker, Stuart Cunningham, and Xu Shao, "An audio-visual corpus for speech perception and automatic speech recognition," *The Journal of the Acoustical Society of America*, vol. 120, pp. 2421, 2006.
- [10] T. Eerola and J. K. Vuoskoski, "A review of music and emotion studies: Approaches, emotion models and stimuli," *Music Perception*, vol. 30, no. 3, pp. 307–340, 2013.
- [11] M. L. Chanda and D. J. Levitin, "The neurochemistry of music," *Trends in cognitive sciences*, vol. 17, no. 4, pp. 179–193, 2013.
- [12] O. Lartillot, P. Toiviainen, and T. Eerola, "A matlab toolbox for music information retrieval," in *Data analysis, machine learning and applications*, pp. 261–268. Springer, 2008.
- [13] J. Pätynen, *A virtual symphony orchestra for studies on concert hall acoustics*, Ph.D. thesis, 2011.
- [14] M. A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney, "Content-based music information retrieval: Current directions and future challenges," *Proceedings of the IEEE*, vol. 96, no. 4, pp. 668–696, 2008.
- [15] G. Tzanetakis and P. Cook, "Marsyas: A framework for audio analysis," *Organised sound*, vol. 4, no. 3, pp. 169–175, 2000.
- [16] V. Alluri, P. Toiviainen, I. P. Jääskeläinen, E. Glerean, M. Sams, and E. Brattico, "Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm," *Neuroimage*, vol. 59, no. 4, pp. 3677–3689, 2012.
- [17] T. Lokki, J. Pätynen, A. Kuusinen, and S. Tervo, "Disentangling preference ratings of concert hall acoustics using subjective sensory profiles," *The Journal of the Acoustical Society of America*, vol. 132, no. 5, pp. 3148–3161, 2012.